Francisco Rodríguez

Data Scientist | Machine Learning & AI | Data Analytics & Visualization

fjrguezcerezo@gmail.com | LinkedIn | Portfolio | GitHub

Summary

Data Scientist with **a strong foundation in physics and math**, and a practical, analytical mindset for solving complex problems. I specialize in building clean, interpretable solutions using **Python, R, and SQL**, and enjoy transforming raw data into useful tools—whether through **dashboards**, **automation**, or **predictive models**. Curious by nature and always learning, I focus on delivering real value through clear insights and thoughtful **collaboration with both technical and non-technical teams**. **Currently deepening my AWS skills** to support scalable analytics and cloud-based data workflows.

Work Experience

Cabildo of La Palma – Data Scientist   **January 2024 – January 2025**

- Automated weekly processing workflows by optimizing **API** calls from multiple Cabildo data sources related to air quality and environmental metrics. Developed a **Python (Requests, Pandas, NumPy, OpenPyXL)** script hosted in **ArcGIS Online** as a scheduled **Jupyter Notebook**, which collected, processed, and updated statistical tables automatically on a weekly basis. These reports were exported to **Excel** and **shared with non-technical stakeholders** from the Environment and Emergency Departments, who used them to assess CO_2 levels in the volcano-affected area of Puerto Naos. **This system saved approximately 12 hours per month, eliminated manual updates, and supported weekly decision-making by over 6 non-technical staff monitoring volcanic CO_2 levels.**

- **Led and present** in-depth Exploratory Data Analysis (**EDA**) using **Python (Pandas, Matplotlib, Seaborn)** and **R (ggplot2, dplyr) across 5+ diverse datasets** related to mobility, night sky quality, meteorology, and especially air quality in La Palma. Focused the analysis on the volcanic CO_2 emissions in Puerto Naos, **uncovering a Pearson correlation coefficient of ~0.6 between coastal CO_2 levels and specific tidal cycles.** This insight, previously undocumented, suggested a potential mechanism of gas accumulation linked to tidal movement. **As a result, the Emergency Department (composed of 4 people) initiated a deeper review of safety protocols in coastal areas**, incorporating this finding into their habitability assessments post-eruption.

- Designed and implemented **interactive dashboards using Power BI and Tableau** for departments such as Waste Management and Human Resources. **In the HR dashboard, I enabled segmentation of the corporate staff (1000+ employers)** by rank, department, and contribution group. **For Waste Management, my dashboard revealed that the increase in recycling container units corresponded with a rise in collected tons of paper, packaging, and glass**—helping justify operational changes in the department's annual report.

- I also **developed real-time dashboards in ArcGIS Online, including a mobility dashboard** (vehicles and pedestrians) which I designed and implemented in its early stages, **and collaborated on a digital twin of Puerto Naos for monitoring volcanic CO_2 concentration.** These dashboards became key tools: **the Emergency Management Department relies on the CO_2 dashboard for daily tracking of habitability conditions**, while other visualizations addressed internal analytical needs.

- **Standardized and integrated a legacy dataset of over 3 million records into the current data model** used by the Cabildo's meteorological portal. **This involved cleaning inconsistencies in date formats, location names, and data structure, as well as engineering new derived metrics** (such as thermal sensation, based on temperature, wind speed, and humidity). **My work enabled the organization to extend its accessible meteorological archive from 2022 (the launch year of the new web portal) back to 2016, when the first IoT sensors were deployed.** This unified and extended dataset supports long-term climate analysis and enhances continuity in public-facing dashboards and environmental reporting.

- **Proposed a structured system for tagging and managing CO_2 sensor data**, addressing the lack of consistency and traceability in the existing architecture. The solution included a **PostgreSQL-based schema that automated the generation of unique sensor aliases using address initials and linked installation metadata**—such as installer input and ownership details—from a private form to a public-facing measurement table. **Ensuring future sensor deployments were consistently tracked and their data easily accessible in a centralized PostgreSQL database.**

Education

MSc in Data Science and Big Data – Open University of Catalonia (UOC)

- **Master's Thesis:** Developed a **sales forecasting and audience segmentation system** for Filarmonía de Madrid **using real ticketing data**, applying **machine learning** models (Random Forest, CNN's, RNN's, ARIMA), **clustering techniques** (K-Means, DBSCAN), and **RFM analysis**, while integrating external event data and addressing overfitting through cross-validation and hyperparameter tuning.

BSc in Physics – University of La Laguna (ULL)

- **Bachelor's Thesis:** Analyzed the intrinsic properties of the EMIR instrument (GTC, IAC) through image-based photometric calibration, using **Python scripts** to extract parameters such as linearity, gain, and sensor drift under varying conditions, and **documenting results** following scientific standards **with LaTeX**.

Skills

Programming & Tools: Python (Pandas, NumPy, Scikit-learn, Seaborn), R (ggplot2, dplyr), SQL (PostgreSQL)**,** Git

Data Viz & Cloud: Tableau, Power BI, ArcGIS Online, Airflow, Docker, Excel, AWS (in progress)

Machine Learning & AI: Time Series (ARIMA), Deep Learning (CNN, RNN – TensorFlow, PyTorch), Clustering (K-Means, DBSCAN), Regression & Classification (Random Forest, Logistic Regression), Feature Engineering, Cross-Validation, Hyperparameter Tuning

Languages: Spanish (Native), English (Fluent)